卒業論文　2016 年度（平成 28 年度）

期待和了平均順目の評価による
モンテカルロ木探索の麻雀への適用

慶應義塾大学 環境情報学部

細田 航星

徳田・村井・楠本・中村・高汐・バンミーター・植原・三次・中澤・武田
合同研究プロジェクト

2017 年 1 月

卒業論文　2016 年度（平成 28 年度）

期待和了平均順目の評価による
モンテカルロ木探索の麻雀への適用

論文要旨

近年、2 人零和確定完全情報ゲームであるチェス、オセロ、将棋といったゲームでは人間のトッププレイヤと同等かそれ以上の実力を持つコンピュータプレイヤが提案されている。一方で多人数不完全情報ゲームである麻雀では未だそれに及ばない状態である。これは従来のゲーム木の探索手法を適用するのが難しいためである。本論文では、探索空間が大きくなることでうまく適用できなかったモンテカルロ法について、期待和了平均順目の数理的な評価を利用することで探索空間を小さくし、適用する手法を提案する。この手法の効果を確かめる実験として、多人数性を排除した 1 人麻雀における和了率の比較を、人間のプレイヤや先行研究 AI と行った。またそれをそのまま 4 人麻雀で対戦させた場合の比較も行う。解析の結果、本手法で構築した AI は 1 人麻雀において、平均レベルのプレイヤの和了率を出すことができることがわかった。また、4 人麻雀に適用した場合、シャンテン数を下げるように打つ AI に比べて高いレートを出すことが可能であることがわかった。

キーワード

麻雀, モンテカルロ法, ゲーム情報学

慶應義塾大学 環境情報学部

細田 航星

Abstract Of Bachelor's Thesis   Academic Year 2016

Summary

In recent years, computer games that are equal to or better than the top players of humans have been proposed for games such as chess, Othello, and shogi, which are two-person zero settled perfect information games ing. On the other hand, Mahjong which is a multiplayer incomplete information game has not yet reached it. This is because it is difficult to apply conventional search methods of game trees. In this paper, we propose a method to apply by restricting mathematical evaluation and applied parts of the Monte Carlo method which could not be applied successfully by increasing the search space. For evaluation, comparison of the rate of winning in one mahjong excluding multiplayer was done with human players and previous studies AI. We also make comparisons when we played it with 4 mahjong as it is. As a result of the analysis, we found that the AI constructed by this method has comparable ability to that of the average level player.

Keywords

Bachelor of Arts in Environmental Information
Keio University

Wataru Hosoda

目 次

図 目 次

表 目 次

第1章　序論

1.1　本論文の背景

近年、2人零和確定完全情報ゲームであるチェス、オセロ、将棋といったゲームではAIの研究がめざましく、人間のトッププレイヤと同等以上の成績を記録している。一方で多人数不完全情報ゲームである麻雀では未だそれに及ばない状態である。これは従来のゲーム木の探索手法を適用するのが難しいためである。多人数ゲームでは状況によってプレイヤの行動の目的が違うといった難しさがある。この難しさを解消するために、先行研究 [10][9] でも多数行われている通り、相手を考慮しない 1人麻雀プレイヤに対して和了率を求める評価を採用する。麻雀を行う時の人間の思考は1 人麻雀をもとにしているため、1人麻雀から4人麻雀への拡張は容易だと考えられる。

1.2　本論文が着目する課題

1人麻雀においてある局面からゲーム木を展開しようとした場合、次にどの種類牌をツモるかはわからないため、ランダムに決定しノードを展開していくことになる。これを繰り返し行うと探索空間が大きくなりすぎるため、手の選択が難しくなる。これを対処するために、あらゆる手法が提案されている。このような場合に木を展開せずに有望な手を選択する手法として、Upper Confidence Bound (UCB) [1] がある。UCB ではある局面から考えられる全ての手に対して, よい結果 を返しそうな手を重視しつつ、何度もプレイアウトと呼ば れるランダムシミュレーションを行うことで 最善の手を決定する手法である。UCB は局面ごとにこの探索を実行するが, 各局面を完全 に別の局面として扱うため他の局面の探索において得た情 報を利用することができない。この問題を解決するために Linear UCB (LinUCB) [2] という手法が提案されている。これは、局面を特徴で表す ことで、対象とする局面が異なっても, それまで対象とした局面の情報を利用できるといったものである。これらは麻雀について適用した例が報告されている [10] が、いずれの方法も平均プレイヤーに満たない成績となった。

本研究では、モンテカルロ法の問題点である、麻雀に適用すると探索空間が大きくなりすぎてうまく適用出来ない問題に対して、探索空間を小さくするために、2つの解決策を行う。一つはモンテカルロ法のシミュレーションを直接牌姿に当てはめるのではなく、牌姿ごとの局面で数理的に比較できる内容を数理計算を用い削減する方法である。もう一つは、三向聴以下に対してはシャンテン数を下げるように打ち、モンテカルロ法を用いる部分を限定する方法である。麻雀の牌理に関しては、上がりに近づく選択ほど重要である [11] ため、聴牌に近い打牌選択においてモンテカルロ法を適用することで、成績に影響を大きく与える部分でモンテカルロ法の効果が発揮できることが期待される。

1.3 本論文の目的

本研究では和了率を理論値計算によって評価することと和了に近い段階へのモンテカルロ法の部分適用によって、麻雀に対してモンテカルロ法をそのままうまく適用できない問題に対して解決策を提案する。多人数ゲームでは状況によってプレイヤの行動の目的が違うといった難しさがある。この難しさを解消するために、先行研究 [1,2] でも多数行われている通り、相手を考慮しない 1 人麻雀プレイヤに対して和了率を求める評価を採用する。従来のモンテカルロ法を麻雀に適用した例では平均プレイヤーレベルの実力が出ていないため、本手法によってそれらを上回ることが期待される。

1.4 本論文の構成

本論文の構成について述べる。2 章では、不完全情報ゲームの関連研究について述べる。3 章では、1 人麻雀にモンテカルロ法を適用する際に、適用範囲を限定し、比較内容を数理的和了率によって評価する手法を提案する。5 章では、3 章で提案した手法で実装した AI に 1 人麻雀を打たせ、各種パラメータを比較する。また、4 人麻雀で打った際の成績の評価も行う。6 章では、本手法によって得られた知見について述べる。

第2章　背景

本章では不完全情報ゲームの関連研究を述べ、麻雀の関連研究がどのように行われているかを述べる。

2.1　不完全多人数情報ゲームの例

プレイ人数に着目した研究としてはポーカーを用いたもの [4],[5] がある。これらの手法では 2 プレイヤでのゲームを強くしてから多人数に適応する方法がとられている。ポーカーは 2 プレイヤであればナッシュ均衡戦略を用いることで世界チャンピオンに勝っているが、ポーカーは 2 人から 10 人程度で参加可能なゲームであり人数が増えると状態数が指数関数的に増大するためナッシュ均衡戦略の計算は難しい。そこで 2 人で行われたナッシュ均衡戦略を 3 人限定で拡張する方法、プレイヤの行動を削減、抽象化することでより少ない人数の少ないゲームを想定する方法がとられた。

2.2　麻雀の例

麻雀においての AI 研究については機械学習を用いた水木ら [6] や北川ら [7] の研究が報告されているが、いずれも平均プレイヤの実力に至っていない。一方、水上ら [?] は 1 人麻雀プレイ ヤの学習に牌譜との一致を目指した平均化パーセプトロンを用い、平均プレイヤ以上の実績を収めている。

水上らの研究より、麻雀においてもプレイヤの人数を削減してから多人数に適用する方法は有効である可能性が高く、以後ゲームを単純化した 1 人麻雀に対しての研究が報告されている。

2.3　麻雀についてモンテカルロ法を応用した研究

2.3.1　モンテカルロ法

水上ら [?] の研究の前衛である、「麻雀を打つ機械」[3] の牌効率のアルゴリズムでは、モンテカルロ木探索を用いている。そのアルゴリズムの概要を図??に示す。ツモを入れた状態で１４枚の牌姿から、それぞれ切ることのできる牌を切った場合を一つのノードとする。この場合深さ 1 の１４つのノードができることになる。この後、それぞれのノードの状態から、ツモることの可能な牌をランダムにツモっていくシミュレーションを行う。この場合ツモる事のできる牌とは、場や

自分の手牌に情報として見えている牌をゲーム開始時に存在する麻雀牌セットの中から抜いたものである。通常麻雀のルールでは、牌をツモることで次に捨てるという動作が発生するが、ここではそれを行わない。これをしないことで、探索空間が膨大に膨れ上がるの防ぐ効果があり、特定のシミュレーション時に最速で上がるための手順を特定することが可能となる。ツモを連続で行っていき、それまで手牌に残っていた牌の組み合わせで和了することができる場合となるまでシミュレーションを行い、上がった場合にプレイアウトとなる。このシミュレーションを各ノードでそれぞれ十分な回数行い、プレイアウト発生するまでのツモった回数の平均とり評価値とする。最後に、各ノードの評価値を比較し、最もこの値が少なかったノードが最適な打牌と考え選択することになる。プレイアウトまでにツモった回数の平均が少なければ少ないほど、より早く和了できることが期待されるからである。

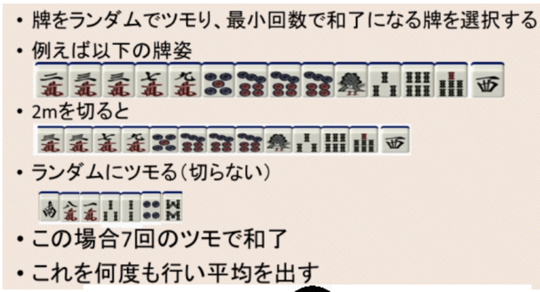

図 2.1: ム

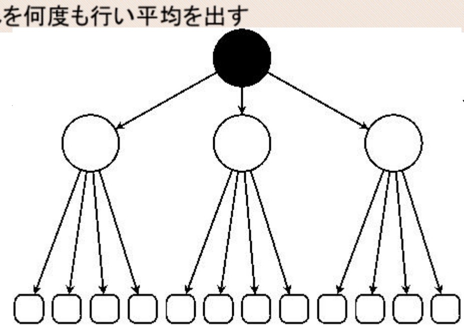

図 2.2: モンテカルロ木探索

2.3.2 UCB1

以上で述べたモンテカルロ法の問題点は、各ノードに対して行うシミュレーションがそれぞれ膨大なため、プレイアウトまでの回数が多く精度が落ちてしまうことである。この問題を解決するために、提案されている手法の中に、Upper Confidence Bound (UCB) [1] を用いたものがある。UCB ではある局面から考えられる全ての手に対して, よい結果 を返しそうな手を重視しつつ、何度もプレイアウトと呼ば れるランダムシミュレーションを行うことで 最善の手を決定する手法である。有望な手に対して多くのプレイアウトを実行することで、先に述べた通常のモンテカルロ木探索よりも精度を上げることができる。以下、このアルゴリズムを詳しく説明する。

通常のモンテカルロ木探索では、ノードごとに期待される手の良さを途中で判定していないため、全てのノードに対して同じ回数のプレイアウトを行うことになる。したがって、良い手が期待できないノードに対しても多くのプレイアウトを割り当てることになり、無駄なシミュレーショ

ンを行ってしまう事が考えられる。また、ノードによっては手の良さを決定する分岐がプレイアウトに対して大量にある場合、他のノードと同じ回数のプレイアウトでは十分な評価が出来ない可能性もある。このような問題に対して、UCB では、UCB 1 値を用いることによって各ノードの有望さをその都度確認し、全体のプレイアウトの回数や特定のノードで行われたプレイアウトの回数を考慮して、プレイアウトをどのノードに数多く割り当てるかということを判別することを考える。UCB 1 値は、x_j は子ノード j の平均報酬, α は定数,n は親ノードの探索回数, n_j は子ノード j とした時、

$$UCB1 = \overline{x_j} + \alpha\sqrt{\frac{2\log n}{n_j}} \tag{2.1}$$

で表される。式 2.1 の右辺の第 1 項は平均報酬を, 第 2 項は信頼度を示している。従来のモンテカルロ木探索では第 1 項の平均報酬をそのまま全てのプレイアウトが終了するまで行った値として利用していたが、UCB1 ではここで第 2 項の信頼度を用いることで全てのプレイアウトが終了するよりも前に信頼できる範囲を推定する。そうすることで、多くのプレイアウトを最後まで割り当てなくとも、一定の信頼度で少ないプレイアウトでそのノードが有望かどうかを判別することが可能になる。信頼度はそのノードのプレイアウト回数が少ないと大きく, 多いと小さくなる。UCB1 値を用いることで レーションを行う事 ができる。

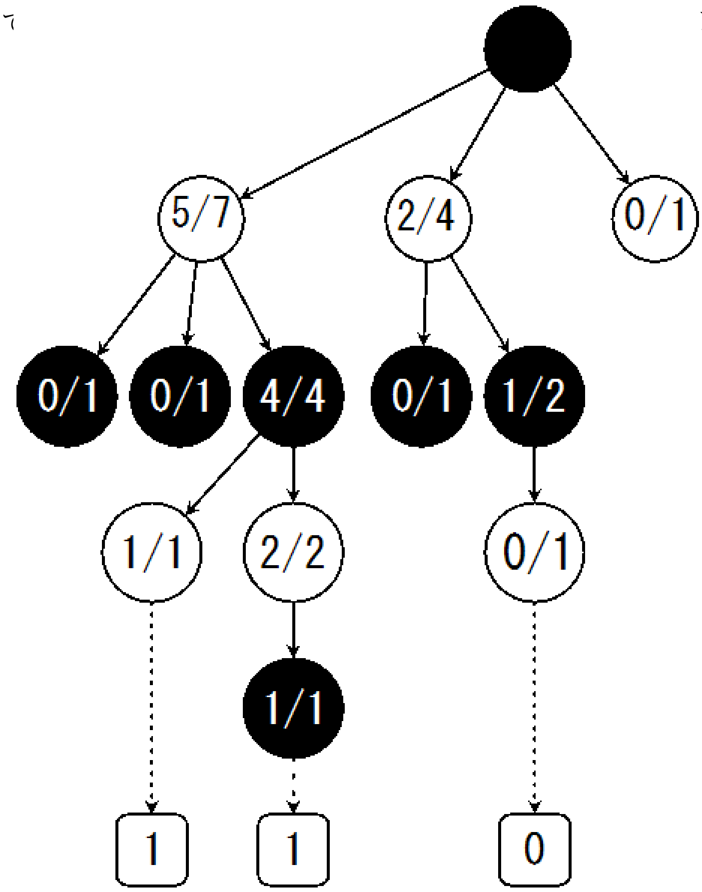

図 2.3: UCB を用いたモンテカルロ木探索

2.3.3　Lin UCB

UCB1 は各局面ごとにこの手順を実行するが, 各局面を完全に異なる局面として扱うため, 他の局面の探 索において得られた情報を共有することができないという 欠点を持つ. この欠点を補うアルゴリズムとして期待され ているのが LinUCB である.

UCB は局面ごとにこの探索を実行するが, 各局面を完全 に別の局面として扱うため他の局面の探索において得た情 報を利用することができない。この問題を解決するために Linear UCB (LinUCB) [2] という手法が提案されている。これは、局面を特徴で表す ことで、対象とする局面が異なっても, それまで対象とした局面の情報を利用できるといったものである。

LinUCB [5] は, UCB を局面を特徴で表すことができる ように拡張したものであり, 牌譜の局面からの教師あり学 習や異なる探索の結果の共有ができる手法である。LinUCB はプレイアウトを行う子ノードを選択する評価値の計算に, 重みベクトル, 特徴ベクトル, 特徴の頻度を表す相関行列を 用いる. 計算により求めた評価値が最大となる子ノードに 対してプレイアウトを行い, 共通で保持する重みベクトル を更新する. 重みベクトルは, 選択したノードの特徴ベクト ルの各項にプレイアウト結果の報酬を乗じた値を, 重みベ クトルの対応する項に足し込んで更新する. この更新によ り, 高い報酬を得た特徴は大きな重みを, 低い報酬の特徴は 小さな重みを持つようになるため, 当該ノードだけでなく, 同様の特徴を持つ他のノードの評価値も更新される. これ により異なる局面で得られた情報を共有し, 利用すること ができる. また LinUCB は探索を行った結果を重みベクト ルとして記録しておくことで, 事前学習の結果を用いる探 索として利用することも可能である. LinUCB のアルゴリズムを Algorithm 1 に示す. ただし xt,at はプレイアウト回数が t 回目の時の選択肢 a の特徴ベ クトル, A は特徴の共起を含めた頻度を表す相関行列, b は各ノードごとの報酬の総和を表すベクトル, θ t は重みベク トル, pt,a は選択肢 a の評価値, α はバランスパラメータ, rt は報酬を表している. rt はプレイアウトにより与えるか, 学習データにより与えるものとする. LinUCB は 4 行目で 前回のプレイアウト結果を反映して重みベクトル θ t を更 新する. 7 行目で重みベクトル, 特徴ベクトル, 相関行列を 用いて各ノードの評価値を計算し, 9 行目で評価値が最大 となるノードを選択する. 10 行目でプレイアウトを行い報 酬を受け取り, 11 行目, 12 行目で A と b の更新を行う. LinUCB は Algorithm 1 の 7 行目に示したように, 式 (2) により各ノードの評価値を求める. 式 (2) の第 1 項は各ノー ドの平均報酬を計算しており, 第 2 項で信頼度を計算して いる.

$$pt, a = \theta\ Ttxt, a + \alpha \tag{2.2}$$

また, UCB で用いている評価値も, 平均報酬と信頼度の和 で計算される. つまり LinUCB と UCB は本質的に等しい 計算をしていることが分かる. また, LinUCB の第 2 項は データ数の増加により十分速く小さくなることが保証され ている [5]. 以上より, LinUCB は UCB と近い運用を行うことができると考えられる.

これらは麻雀について適用した例が報告されている [10] が、いずれの方法も平均プレイヤーに満たない成績となった。

2.4 本論文が着目する課題

1 人麻雀においてある局面からゲーム木を展開しようとした場合、次にどの種類牌をツモるかはわからないため、ランダムに決定しノードを展開していくことになる。これを繰り返し行うと探索空間が大きくなりすぎるため、手の選択が難しくなる。これを対処するために、あらゆる手法が提案されている。このような場合に木を展開せずに有望な手を選択する手法として、Upper Confidence Bound (UCB) [1] がある。UCB ではある局面から考えられる全ての手に対して, よい結果 を返しそうな手を重視しつつ、何度もプレイアウトと呼ば れるランダムシミュレーションを行うことで 最善の手を決定する手法である。UCB は局面ごとにこの探索を実行するが, 各局面を完全 に別の局面として扱うため他の局面の探索において得た情 報を利用することができない。この問題を解決するために Linear UCB (LinUCB) [2] という手法が提案されている。これは、局面を特徴で表す ことで、対象とする局面が異なっても, それまで対象とした局面の情報を利用できるといったものである。以上のような対策が考えられてはいるが、1 人麻雀について適用した結果平均プレイヤーの実力までに至っていない。囲碁においては一定の成果を上げたものの、麻雀ではうまく適用できない問題に対しては、麻雀の場合はより探索の空間が大きく、UCB 1 値や LinUCB を指標とするノードの展開の方法ではうまく適用できていない可能性がある。したがって本研究では、ノードの展開を行う際に指標とするものを麻雀というゲームの性質上の状況のパラメーターを用意し、それを使うこととする。

第 3 章　提案手法

本章では、以上で述べた課題に対して本研究が提案する手法を述べる。

3.1　手法の概要

本研究が提案する手法は、期待和了巡目という途中局面の静的指数を評価することで、適切な打牌を選択するアルゴリズムである。与えられた牌姿において、和了形までにかかるツモの回数を和了までの消費巡目と定義する。ここで、麻雀においてはどのような牌をツモってくるかは確率的にしかわからず、確定的な情報ではない。したがって、和了までの消費巡目は同じ牌姿であったとしてもそれぞれの場合において異なる可能性がある。そのため、それらを十分な回数行った場合に収束することが期待される平均的な消費巡目を考える。これを、この論文では以下その牌姿における期待和了巡目とする。

図 3.1: 期待和了巡目

3.2 期待和了巡目の算出

3.2.1 変化を考慮した期待和了巡目の計算

国士氏の研究により、ある手牌における聴牌までの平均消費順目は、それぞれの変化での平均消費巡目のそれぞれの変化する確率での単純な平均で与えられることがわかっている。[3] すなわち、図??に示すように、元の成功率が p、手変わりする確率が q,r、手変わり後の成功率が Q,R ならば、

その手の向聴が進むまでの平均消費巡目は

$$\frac{p\frac{1}{p}+q\frac{1}{Q}+r\frac{1}{R}}{p+q+r} \tag{3.1}$$

と表される。

変化の牌姿

この式 3.1 は、他プレイヤによって捨てられた牌を使ったシャンテン数現象パターン、すなわち鳴きを考慮していない。したがって面前聴牌のみを考えた式ということになる。しかし、1 人麻雀においては相手プレイヤを考慮しないため、1 人麻雀における聴牌率は 4 人麻雀における面前聴牌確率と一致し、この式によって求まる。鳴きによって聴牌する有効牌については、全て面前ツモで聴牌する有効牌の一部であり、制限によって鳴くことが出来ない有効牌も多く存在する。例えば、以下のような牌姿の場合である。

ヘッドレスイーシャンテンの牌姿を見て解説図

また、面前聴牌と鳴きによる聴牌の違いについては、麻雀の役の性質上、和了することができなくなる制約や、点数が下がるなどの問題がある。したがって、鳴くことによる聴牌が面前聴牌に与える影響は少ないと考えられる。

ここで、本論文ではこれを和了時までの式に拡張する。

まず平均の関係から

$$\frac{p\frac{1}{p}+q\frac{1}{Q}+r\frac{1}{R}}{p+q+r} \tag{3.2}$$

ここで和了期待順目について考える

前節と同様に、4 人麻雀において聴牌した

面前で聴牌するための平均消費順目と鳴きを考慮

まず、聴牌までの平均順目と聴牌後の和了までの平均順目の合計がすなわち全体の平均消費順目であることを証明する。次に、それを合成した結果の数式がどうなるかを書く。

考察の結果このようになる。

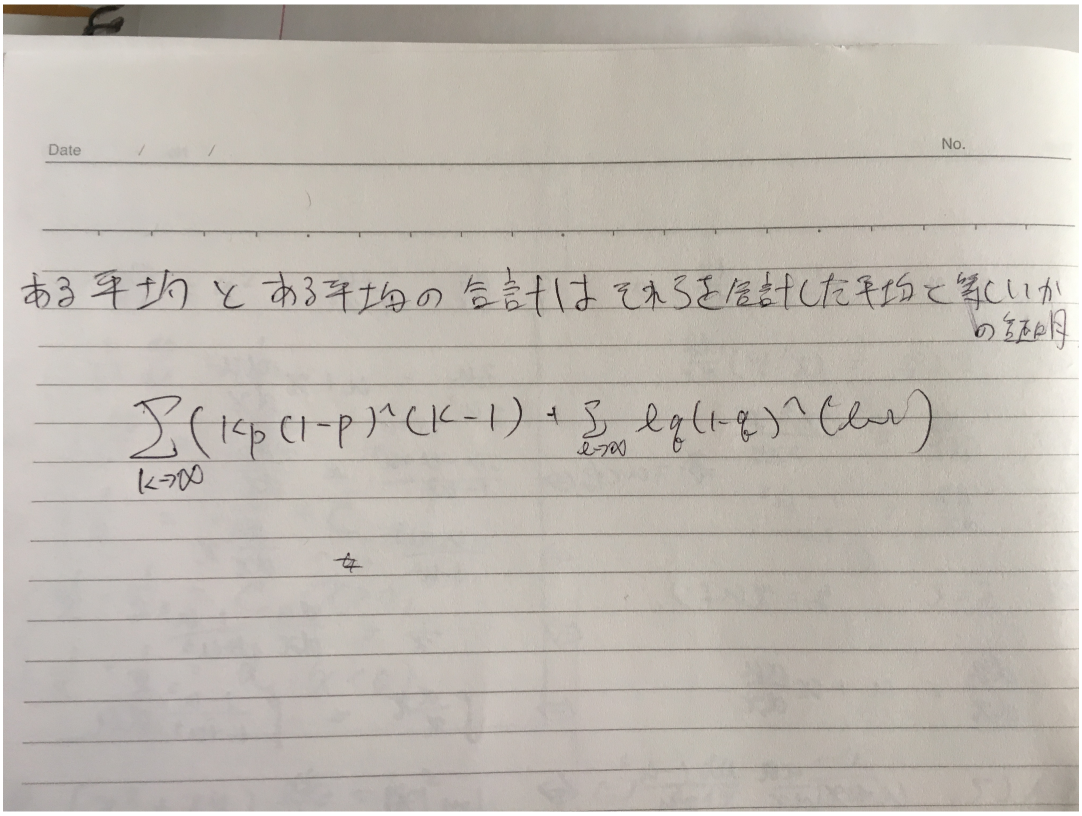

図 3.2:

$$\frac{p_{\frac{1}{p}} + q_{\frac{1}{Q}} + r_{\frac{1}{R}}}{p+q+r} \quad + \text{和了期待順目} \tag{3.3}$$

また、(現在の平均テンパイ巡目)=(次巡の平均テンパイ巡目)+1　であることから、各ノードが手替わり率とテンパイ率を値として持つ深さ 1 の木構造で記述できる。したがって、同様に、n 手先の手替わりまで考えると深さ n の木構造で記述することができる。n を大きくすることにより、精度の高い和了率を近似することが可能となる。

本研究では、この式 (3) をそれぞれの牌姿に当てはめ、その計算結果が高くなるようなノードを探索する。これにより、従来の手法では何回も同じ牌姿の変化をシミュレートしなければいけない問題があったが、(3) 式の評価によってその多くが削減できる。したがって、この方法によって精度が高くなることが期待される。

3.3　想定されるメリット

この節では、以上に提案した手法を関連研究の事例を交えて、優位性があると期待される部分について述べる。まず最初に本手法と同じく与えられた牌姿に対して途中局面の静的指数を評価することで妥当である牌姿を選択するアルゴリズムを挙げる。これらは、モンテカルロ法などのシミュレーションをプレイアウトまで行わないものである。また、関連研究としては本手法と同

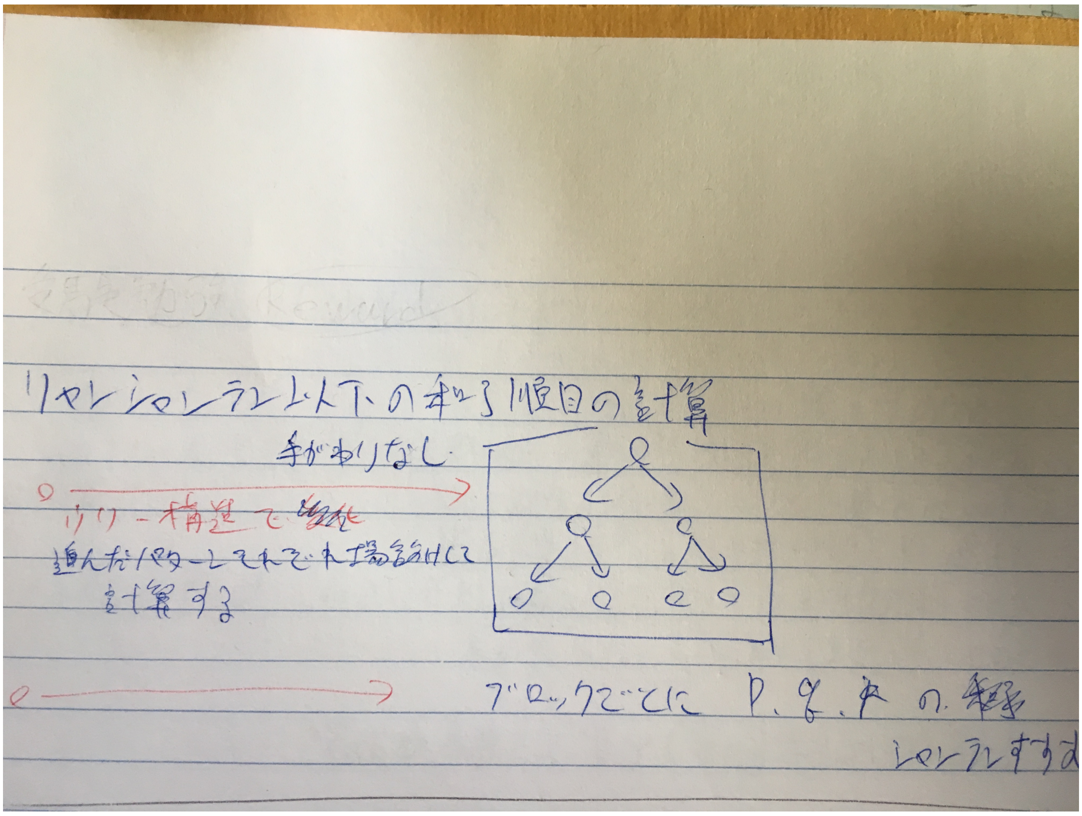

図 3.3:

じく期待和了巡目と近い考え方を用いた例を述べる。しかし、この例では期待和了巡目の算出にモンテカルロ法によるシミュレーションを使っており、静的な評価方法ではない。また、期待和了巡目の算出方法も、ツモったあとに切らないという考え方を使っており、正確な評価ではない点がある。

3.3.1 シャンテン数を下げるアルゴリズムとの比較

麻雀において、特定の和了までに必要な牌の数がシャンテン数であるため、与えられた牌姿においてのシャンテン数を把握することは和了のしやすさを評価することにつながる。これは最も基本的な方法であるため、あらゆる麻雀 AI の研究で基礎として行われている。三木ら??は、UCT アルゴリズムの評価を行う際に、グリーディングプレイヤーとして、シャンテン数を下げるように打つプレイヤーを用いている。このグリーディングプレイヤーでは、与えられた牌姿の中で打つことのできる全ての牌に対して、打った後のシャンテン数を調べる。その後、元の牌姿とそれらのシャンテン数をそれぞれ比較し、シャンテン数が小さくなるような手を選択するアルゴリズムである。シャンテン数が同じように小さくなる手が複数あった場合には、ランダムでその中から選んだものを選択するようになっている。

この手法の問題点は～この代わりとしては有効牌を数えるという手が挙げられる

図 3.4: シャンテン数下げるように打つアルゴリズム

3.3.2　有効牌の数を数えるアルゴリズムとの比較

・有効牌の方法・数え上げ論文の紹介

図 3.5: シャンテン数下げるように打つアルゴリズム

この手法の問題点

3.3.3　擬似的に期待和了巡目をモンテカルロ法で行った研究との比較

期待和了巡目シミュレーション

3.4 想定されるデメリット

第4章　設計と実装

4.1　実験概要

4.2　１人麻雀プレイヤーの設計と実装

提案手法で提示した、期待和了平均順目の評価によるモンテカルロ木探索を評価するために、１人麻雀プレイヤーを実装した。１人麻雀プレイヤーとは、相手プレイヤーを考えない多人数性を排除した麻雀のことである。ルールについては次の節で詳しく説明している。本研究で実装した 1 人麻雀プレイヤーのフローチャート図を図 4.1 に示す。まず、１人麻雀プレイヤーにおいて開局から、終局まで行う対局のループを定義する。開局時には、山と配牌を設置する。その後山から一つ牌をツモり、一つ捨てる動作を繰り返す。終了条件を満たした時点で終局となり、その局の成績を集計する。このループを 1 人麻雀プレイヤーにおける 1 対局と定義する。この対局のループを十分な回数行い、和了率を測定する。

4.3　１人麻雀のルール

多人数ゲームでは状況によってプレイヤの行動の目的が違うといった難しさがある。この難しさを解消するために、相手を考慮しない 1 人麻雀プレイヤを考える。この 1 人麻雀プレイヤのルールは、相手プレイヤを考えないため、相手プレイヤーに点数を支払う事による放銃や被ツモ失点などを考えない。また、相手プレイヤーによる捨て牌が存在しないので、鳴きや栄和を考えない。リーチについても、和了率の観点では不要なため、考えない。門前でツモを繰り返し行うだけのシンプルなシステムである。先行研究 [?] 評価を合わせるため、ツモの回数は 27 回とした。

表 4.1: 本研究の 1 人麻雀プレイヤのルール

アクション	4 人麻雀	1 人麻雀
和了	ツモ和了、ロン和了どちらも可	ツモ和了のみ
点数	失点や和了の点数を考慮する	点数は考えない
リーチ	可能	なし
鳴き	チー、ポン、カンが可能	なし
終局	誰かが和了するか、ツモ山 70 枚がなくなるまで（136-14-13*4)	和了するか、27 回ツモるか

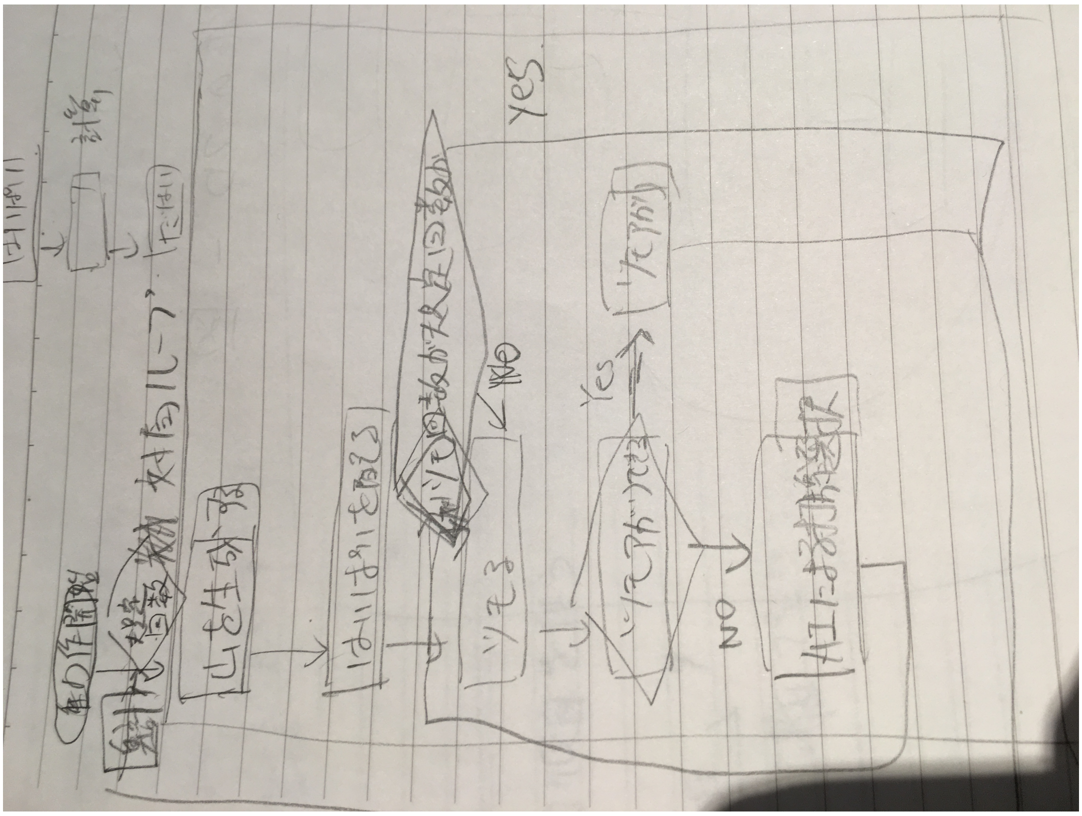

図 4.1: 1 人麻雀のフローチャート

また 1 人麻雀の成績の評価においてはその点数の推移よりも、和了率自体が重要だと考えられている。したがって、点数という複雑なパラメータを省いた和了率に注目したルールとなっている。その理由としては、以下のような理由が述べられる。

まず、4 人麻雀では平均順位の低いプレイヤほ ど、平均和了点低く、和了率が高いことが統計で明ら かになっている [8] したがって強いプレイヤは点数よりも和了率を重視して打っていることがわかる。

上手いプレイヤの平均和了点が低くなる理由として手役を無理に狙った打ち方をしないためである。また麻雀の点数の特性上、満貫までの点数は指数関数的にその得点が増えていくが、満貫以上の場合は線形に近くなる。したがって、難易度に対して望める点数が割に合わない高い和了を狙うより、比較的低い点数で多く和了することが効率が良い。また、失点の観点からも、４人麻雀では自分が放銃をしない場合も他プレイヤーの和了によって被ツモ失点を被ることがある。これに対しても、和了すれば他プレイヤーが和了することができないため、相手のチャンスをつ

ぶすといった意味でも和了率の高さは大事である。

4.3.1　終了条件

・ツモが規定回数かどうか・そもそもなんでツモ回数を２７としてるのか？

4.3.2　全体のループ回数

4.3.3　山のデータセット

同じように用意したもの

4.3.4　AI による打牌選択

・どのような手法を選択したか爆打モンテカルロ法 UCB １モンテカルロ法 LinUCB を用いた方法・どのように実装したか　どのような制限を加えたか（プレイアウトの回数、CPU や実装コード）計算量の問題

4.3.5　シャンテン数の計算方法

バックトラック法、ハッシュ法

4.4　４人麻雀における実対戦用システム

本研究の提案手法のアルゴリズムが４人麻雀でも有用かどうかを調べるために、４人麻雀で実際に対戦するための自動打ちシステムを構築した。このシステムを動かす場所として、オンライン麻雀サイト「天鳳」を選んだ。

・環境・VM 上

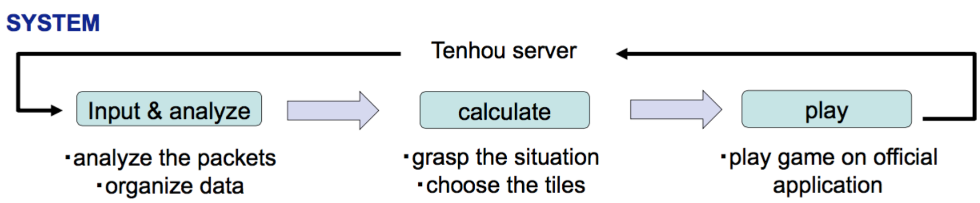

図 4.2: 天鳳上で自動打ちするためのシステムの設計図

4.4.1　オンライン麻雀サイト天鳳

天鳳は、現在インターネット上で麻雀を打つことができるサイトの中で最も利用者が多いサイトであり、登録者は３９０万人を超える大型のサイトである。また、アクティブユーザー (180 日以内の対戦履歴があるプレーヤ数) は２７万人を超え、現在最も活気のあるインターネット上の麻雀フィールドである。天鳳では麻雀研究に対する支援が充実しているという面も大きい。最高ランク者のみが打つことのできる鳳凰卓というクラスの試合データ（以下牌譜）を全て公開しており、より良質な強者の牌譜を取得することができる。これらの牌譜は一日に約５００試合ほど手に入り、量ともに十分利用できる。これについては、筆者も強者の統計的な手順について研究する際によく利用させていただいた。また、天鳳では AI で麻雀を打つことを一定の条件下で許可している点も大きい。鳳凰卓については人間同士の生粋の強者のフィールドというコンセプトがあるため許可されていないが、その下の特上卓以下では条件を揃えることで対戦することが可能である。一番下の一般卓については、最も低ランクなため対戦人数も多いため、AI による影響が少ないと考えられている。そのためある程度対戦が可能な AI であれば、他プレイヤーの迷惑にならないような範囲での利用が許可されている。また、一般卓と鳳凰卓の間の上級卓・特上卓では、AI を用いて他のプレイヤーの不正検出に協力するという条件を元に AI の稼働を許可されている。AI を稼働させることによる他プレイヤーのメリットを提供することで、反対するものが少なくなるからである。本研究のシステムは、一般卓で稼働させた。

図 4.3: ランク別リスト・AI 稼働条件

4.4.2　プレイさせた天鳳のルール

・ルールフロー図‥ 天鳳上の制約（ルール）・天鳳上の制約（マシン上の問題　持ち時間）

4.5　全体フローの詳細説明

4.6　パケット入力（input）

・Kmo2 さんのやつを利用・どのようにパケットを読んでるか？（長く解説する必要はないかもしれない）

4.7　計算（後ほど詳しく説明）

・天鳳からの情報をどういうふうな構造体で保存しているかどうかとか。

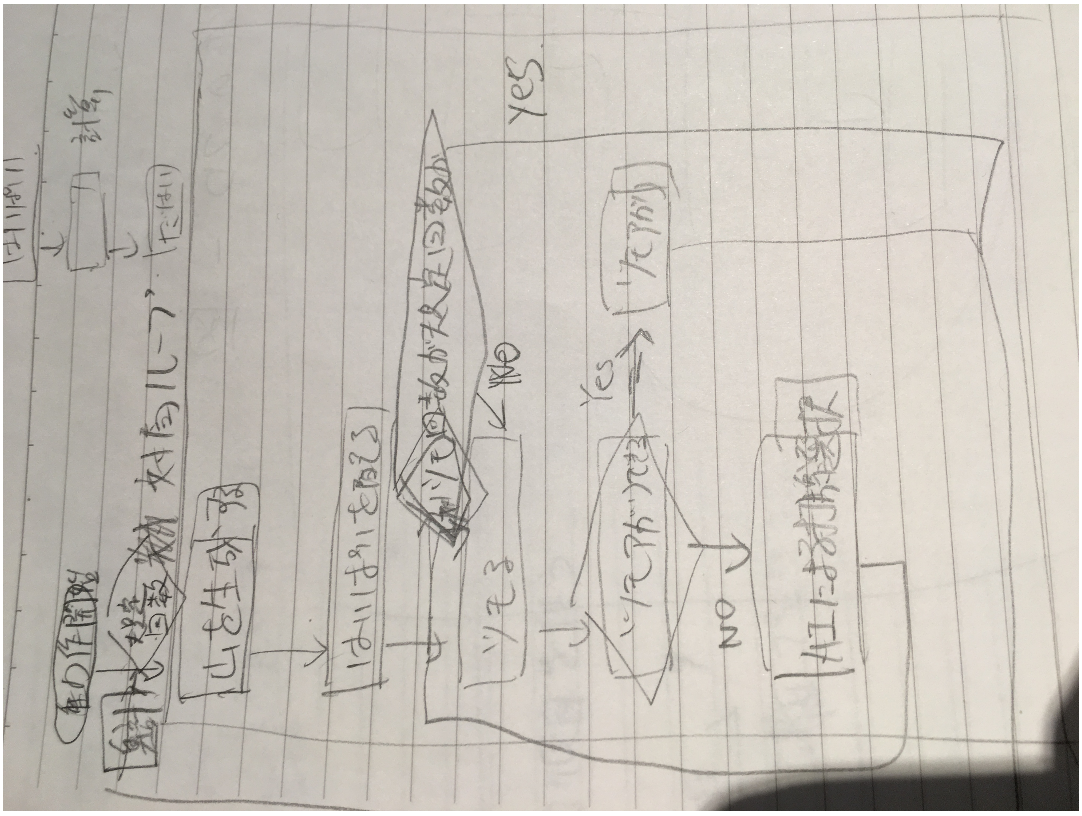

図 4.4: 4 人麻雀自動打ちシステムのフローチャート

4.7.1　シャンテン数の計算方法

バックトラック法、ハッシュ法

4.8　アウトプット

・打牌を選択して、天鳳サーバーにそれを送る・公式クライアントを使うので、マウスハンドラを使う・配列でどの牌を切るかを選択した上で、左から何番目かを考えてクリックさせる・リー牌に対してのアクション

第5章　評価

5.1　１人麻雀による和了率の比較

提案手法で提示した、期待和了順目の評価による打牌選択のアルゴリズムを１人麻雀に適用し、性能を評価した。図 5.1 に示すのは、同じ５０００局のテストデータを与え、対局を行うごとに和了率を測定し更新していったときのグラフである。また、天鳳において実力が上位０．１％に当たる上級者（本論文執筆者）と、天鳳において全体の５０％に当たる平均プレイヤーを一人ずつ用意し、１００局のテストデータで同じく和了率を比較した。5.1 この場合も同じテストデータで１００局他のアルゴリズムによる和了率も測定している。図に示す「シャンテン数」のアルゴリズムとは、シャンテン数が最も少なくなるような牌を選択して打牌するアルゴリズムで、そのような牌が複数存在する場合はその中から切る。「有効牌」のアルゴリズムは、「シャンテン数」のアルゴリズムの中で、複数の牌が存在するときにその有効牌の枚数を比較して多いものを切るアルゴリズムである。最後に、「期待和了巡目」とあるアルゴリズムは、本研究で提案した、期待和了巡目の最も小さいものを選択するアルゴリズムである。

表 5.1: １００局のデータセットでの和了率

手法	あがった局数（%)
上級者	53
期待和了巡目	44
有効牌	41
平均プレイヤ	36
シャンテン数	32

１人麻雀を５０００局打たせたことによる和了率の比較では、「シャンテン数」＜「有効牌」＜「期待和了巡目」という結果となった。これは統計的に優位な水準での優劣である。結果については期待通りで、シャンテン数を下げるだけのアルゴリズムに対し、それをさらに有効牌の数を比べるアルゴリズムでは和了率が上昇した。また、本研究で提案した「期待和了巡目」のアルゴリズムでは、さらに有効牌の中でもより和了までの到達度を正確に計算することで和了率が上がることが確認された。１人麻雀による対局では、多人数性が存在しないため、本手法がうまく適用できると考えられる。また、上級者と平均実力者の人間プレイヤーを加えた１００局の対局では、期待和了巡目のアルゴリズムの和了率は平均プレイヤーより優り、上級者に劣る結果となった。これは、上級者の場合はシャンテン数が瞬間的にあえて最小でないような打牌をして結果的

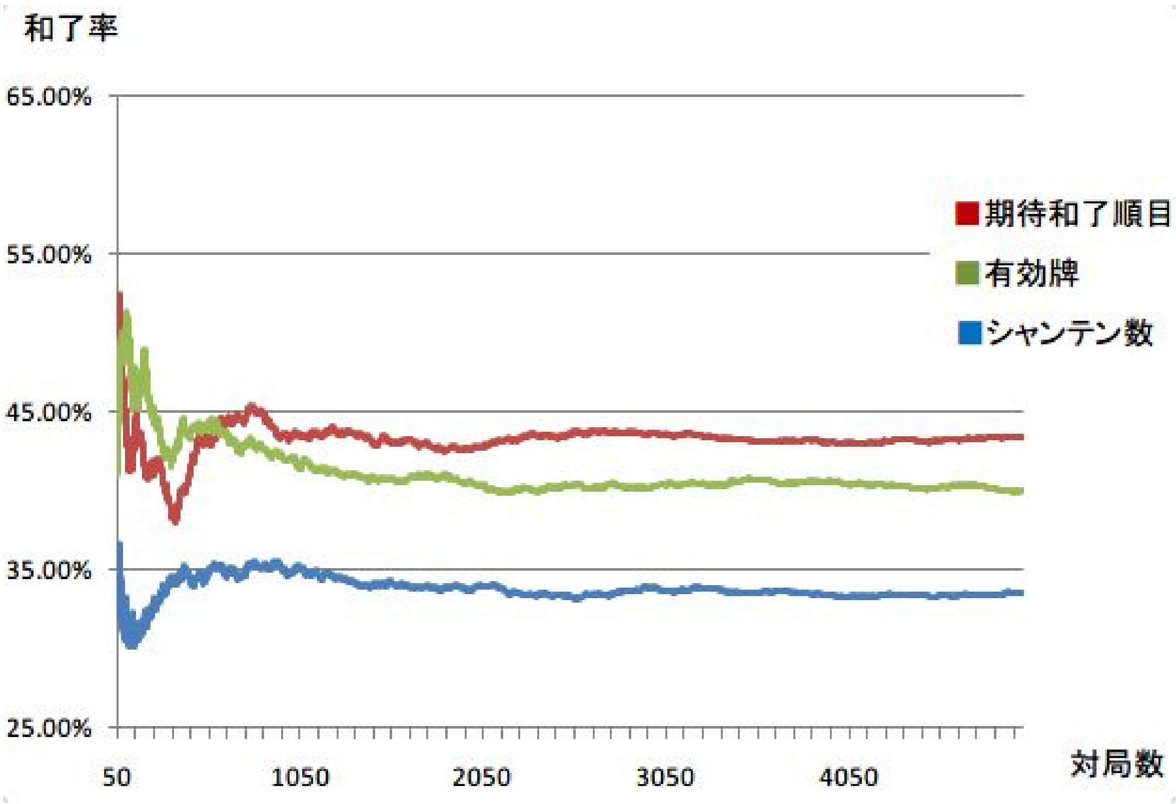

図 5.1: １人麻雀の和了率遷移

に和了率がもっとも高くなる場合が存在するからであると考えられる。期待和了巡目はシャンテン数が最小になる牌の中から打牌を選択しているため、このような選択が行えない。

5.2　４人麻雀における成績の評価

本研究では１人麻雀における和了率の上昇を目指すため、静的指数である期待和了巡目を利用し、打牌を決定するアルゴリズムを提案した。これを第４章で設計した実装を元に、オンライン麻雀天鳳で打たせ、その成績を集計した。対戦した場所は天鳳の一般卓で、ルールは喰いアリ赤アリの東風戦、持ち時間は３秒である。２０１６年１１月から２０１７年１月までの期間対戦を行い、試合数は２２３１戦となった。同じように、関連研究では１人麻雀のアルゴリズムを４人麻雀に適用し、実際の対人戦でその成績を評価している例が存在する。この節では、それらの研究の成績と本手法の比較を行い、その優位性を評価した。

評価として、和了率、放銃率、レーティングを比較した。

5.2.1　和了率

４人麻雀で実装した自動打ちシステムを対戦させた結果の、和了率の遷移を図 5.2 に示す。５０戦以下の対局では母数が少ないために和了率の偏差が大きいため、図は５０戦以上の対局から

の和了率を掲載した。和了率は３００戦までは偏差が大きかったが、５００戦を超えると徐々に収束していき、２２３１戦後の最終和了率は２１．２９０％となった。

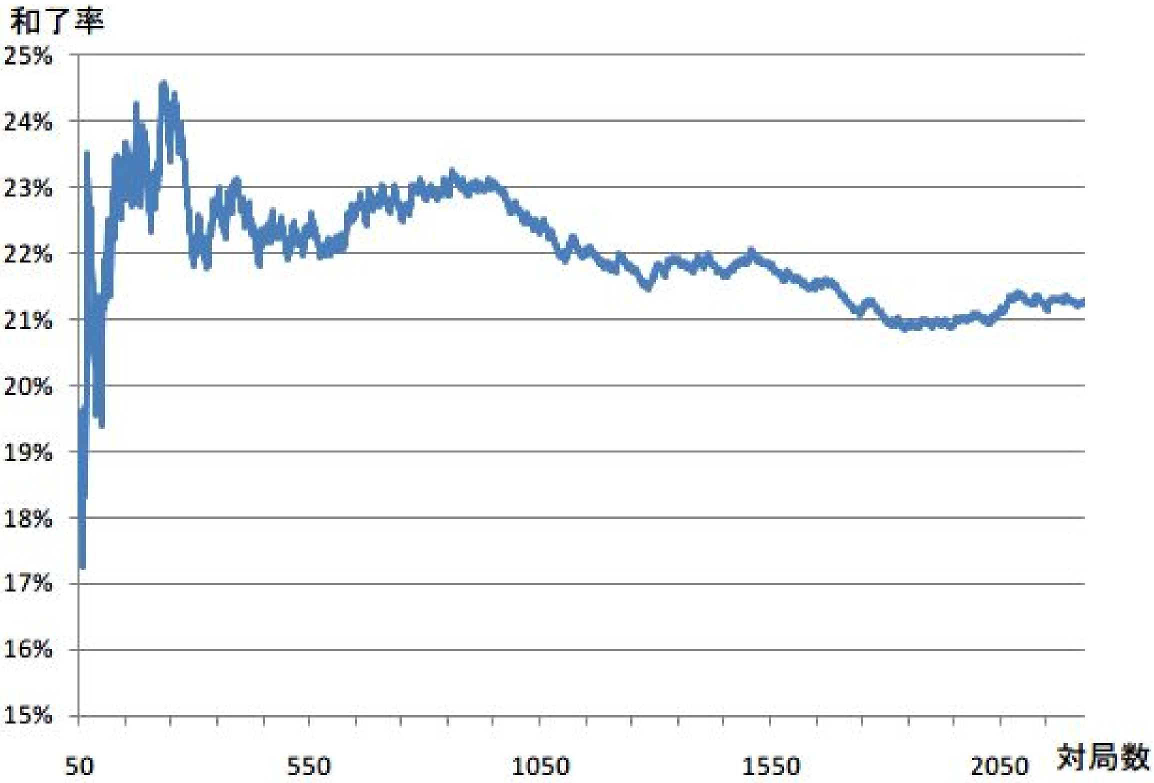

図 5.2: 和了率遷移

また、関連研究と比べた結果の表 5.2 に示す。佐藤らは有効牌を数え上げることによって打牌を選択するアルゴリズムを使用したが、その中で再帰の深さを変えたりヒューリスティックを加えたり複数の手法においてのデータを取っている。ただしそれらによる差は統計的に優位なほど大きな差ではなかったため、今回はそれらの中で最も基本的である再帰の深さ１のものと比較した。また、表にある平均プレイヤーとは、天鳳において平均レートが１５００付近である初段のプレイヤーとした。そのデータは天鳳のランキングページに公開されているものを引用した。

表 5.2: ４人麻雀においての和了率の比較

プレイヤー	本研究	佐藤らの研究	水上らの研究	平均プレイヤー
対局数	2231	2526	504	-
和了率 (%)	21.3	20.1	18.8	21.9

和了率を比較すると、本手法は佐藤らの研究をわずかに上回る結果となった。佐藤らの研究では有効牌の数え上げによって打牌を選択しているが、本手法では平均和了巡目を用いることでより先の展開を考慮した打牌の選択が可能になっていると考えられる。これは事前に期待されていたとおりであった。しかし、大きな差が生まれたというわけではなく、平均プレイヤーを優位に

超えることはできなかった。これは、平均プレイヤーは鳴きによる和了も含まれているためであると考えられる。その理由に、１人麻雀による和了率（鳴きを含まない）は平均プレイヤーを超えていることがあげられる。ただし、４人麻雀でこれ以上の和了率を挙げるためには、やはり鳴きについての和了を加えないと難しいこともわかる。

5.2.2　放銃率

次に、和了率の遷移を図 5.2 に示す。５０戦以下の対局では母数が少ないために放銃率の偏差が大きいため、図は５０戦以上の対局からの宝珠率を掲載した。放銃率は５００戦までは下降を続けたが、１０００戦を超えると徐々に収束していき、２２３１戦後の最終放銃率は１８．３７７％となった。

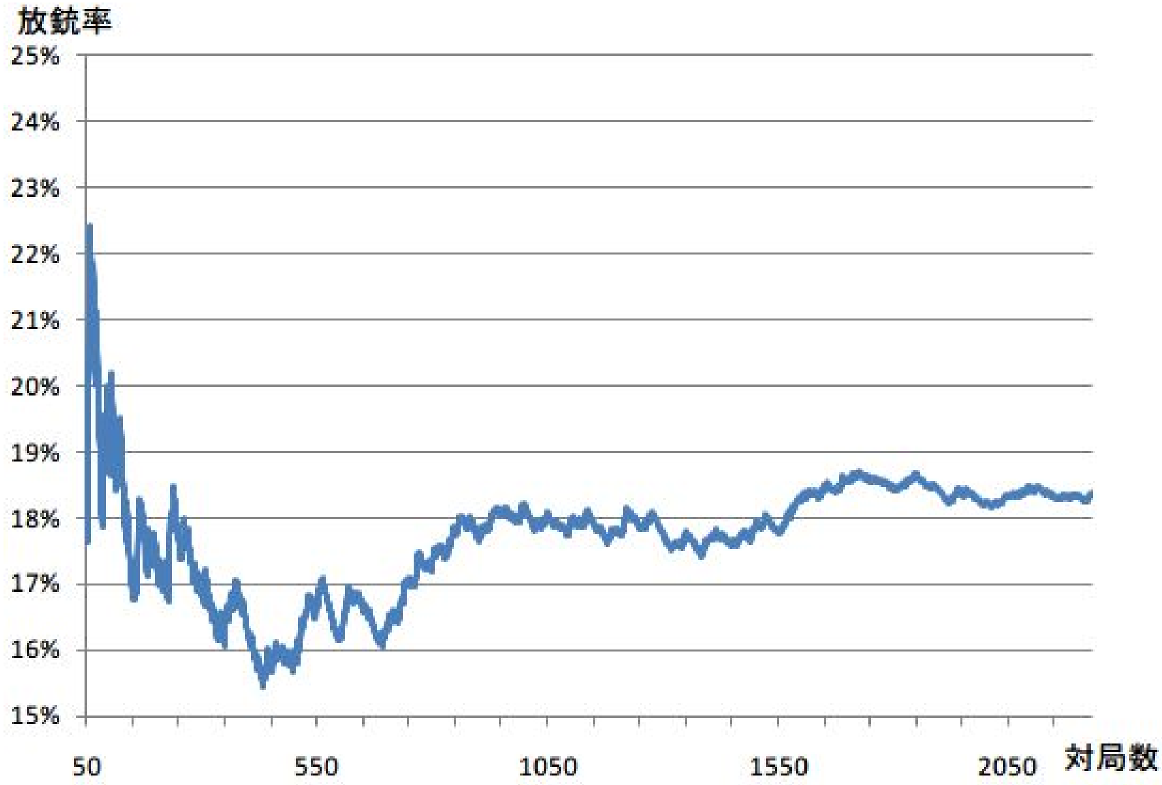

図 5.3: 放銃率遷移

また、関連研究と比べた結果を表 5.3 に示す。放銃率においてはどの研究にも大きな差は見られなかった。これらの研究では、４人麻雀におけるオリの戦略を加えていないからであると考えられる。平均プレイヤーはオリを行っているため、放銃率が低い。４人麻雀においては放銃率による失点は成績を悪くするため、重要な指標であるが、これらはオリの戦略を実装することで改善する必要があると考えられる。

表 5.3: ４人麻雀においての放銃率の比較

プレイヤー	本研究	佐藤らの研究	水上らの研究	平均プレイヤー
対局数	2231	2526	504	-
放銃率 (%)	18.4	18.9	19.0	16.4

5.2.3 レーティング

レーティングとは、特定のプレイヤーが他のプレイヤーと比較してどの程度成績が良いかを図る評価の仕方の一つである。レーティングは、平均順位と負の相関を持ち、式 5.1 で計算される。ゲームを行ったときの卓の平均のレーティングを R_ave とし、ゲームの結果の順位を $Rank$、ゲームを行う前のレーティングを R とした時、そのゲームによって更新されるレーティングが R' である。初期の時点でのレーティング（R）は１５００である。

$$R' = R + (50 - Rank \times\ 20 + \frac{R_ave - R}{40}) \times\ 0.2 \tag{5.1}$$

４人麻雀で実装した自動打ちシステムを対戦させた結果の、レーティングの遷移を図 5.4 に示す。５０戦以下の対局では母数が少ないためにレーティングの偏差が大きいため、図は５０戦以上の対局からのレーティングを掲載した。レーティングは５００戦までは偏差が大きかったが、１０００戦を超えると徐々に収束していき、２２３１戦後の最終放銃率は１３８２となった。

また、関連研究と比べた結果を表 5.4 に示す。レーティングは佐藤らの研究をわずかに上回るという結果になった。しかし大きな有意差は見られず、平均プレイヤーにも及ばなかった。これは和了率の観点では本研究が優位なものの、四人麻雀では多人数性の理由によりほかの影響が多いことが考えられる。理由としては、１人麻雀では和了率が本研究の方が高いものの、放銃率では平均プレイヤーが低く出ているため、オリの影響が大きいと思われるからである。

表 5.4: ４人麻雀においてのレーティングの比較

プレイヤー	本研究	佐藤らの研究	平均プレイヤー
対局数	2231	2526	-
レート	1382	1339	1558

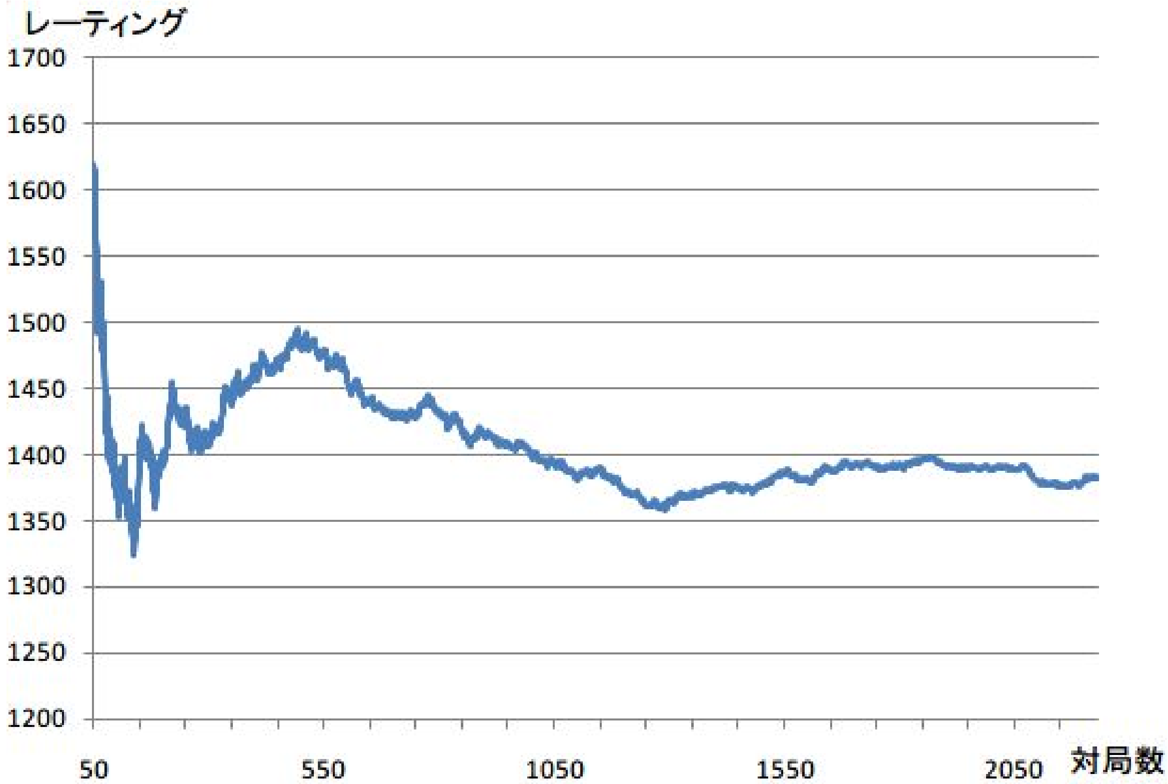

図 5.4: レーティング遷移

第6章　結論

6.1　本研究のまとめ

本研究では、各牌姿における期待和了平均順目の評価をモンテカルロ木探索に利用することによって、モンテカルロ木探索を麻雀に適用することを提案した。評価として、1人麻雀の和了率を比較した。先行研究では UCB 1や LinUCB を1人麻雀に適用する例が報告されていたが、これらは和了までのシミュレーションを行っているものであり、探索空間が多く精度が落ちるため平均プレイヤには及ばなかった。本研究の提案手法では、和了率が平均プレイヤ達することが確認できた。また、期待平均和了順目をどれほど深くの変化まで計算するかに関しては、深く計算することでより精度が保たれることが確認できた。ただし、5巡目以降に関しては大きな違いが見られないことがわかった。4人麻雀の評価については、シャンテン数を下げるように打った先行研究よりもレートが高いということを示すことが出来た。

6.2　本研究の結論

麻雀において従来の探索方法では探索空間が大きくなりすぎて精度が落ちる問題に対して、期待平均和了順目を用いて和了への確率を簡潔に評価することにより、探索空間を小さくする方法が有用であることがわかった。また、期待平均和了順目はおよそ5巡目までの変化を考慮することで十分な精度が期待できることがわかった。最後に、4人麻雀 AI のアルゴリズムとして、モンテカルロ木探索を期待平均和了巡目を指標に行う手法についてはある程度性能を期待できることがわかった。

6.3　今後の課題と展望

今回の検証では、既存のモンテカルロ木探索を行った研究よりも、1人麻雀の和了率は高くなることを示すことが出来た。しかし、牌譜から教師信号を学習した方法により和了率を上回ることは出来ていない。この原因としては、期待平均和了順目を用いる方法では、ツモを無限思考回数行える場合の平均順目を最小にするように考えていることである。実際の麻雀では、ツモ回数が限られていて、相手の和了によってもその終局までの巡目が異なる。今後の課題としては、期待平均和了順目に対して、残りツモ回数の変数を導入したモデルを構築し、同じようにモンテカルロ木探索を行うことである。このようにすることで、より4人麻雀に近い状態での和了率最大化を目指すことができる。

参考文献

[1] Auer, p., cesa-bianchi, n. and fischer, p.: Finite-time analysis of the multiarmed bandit problem, machine learning, vol. 47, no. 2-3, pp. 235256 (2002).

[2] Li, l., chu, w., langford, j. and schapire, r. e.: A contextual-bandit approach to personalized news article recommendation, proceedings of the 19th international conference on world wide web, acm, pp. 661670 (2010).

[3] nmizu, モンテカルロ法を用いた牌効率, コンピュータで鳳凰卓を目指す part4,http://www.nicovideo.jp/watch/sm25236127(2014).

[4] Risk, n. a. and szafron, d.: Using counterfactual re- gret minimization to create competitive multiplayer poker agents, proceedings of the 9th international conference on autonomous agents and multiagent systems: volume 1-volume 1, international foundation for autonomous agents and multiagent systems, pp. 159166 (2010).

[5] 古居敬大:相手の抽象化による多人数ポーカーの戦略の決定, 修士論文, 東京大学 (2013).

[6] 三木理斗:多人数不完全情報ゲームにおける最適行動決定に関する研究, 修士論文, 東京大学 (2010).

[7] 北川竜平:麻雀の牌譜からの打ち手評価関数の学習,proceedings of 12th game programming workshop (2007).

[8] とつげき東北. おしえて!科学する麻雀, 2009.

[9] 水上 直紀ほか. 降りるべき局面の認識による 1 人麻雀プレイヤの 4 人麻雀への適用. 第 18 回ゲームプログラミングワークショップ（GPW2013）, Vol. 31, No. 3, pp. 1–7, 2013.

[10] 中張 遼太郎ほか. Linucb の 1 人麻雀への適用. ゲームプログラミングワークショップ, 2013.

[11] ネマタ. テンパイの技術. 勝つための現代麻雀技術論, 2014.